

Construindo o *Dicionário de valências*: problemas e resultados

Mário A. Perini*

RESUMO

O Projeto VVP (em andamento na UFMG com participantes de outras três Universidades de Minas) tem como objetivo a elaboração de um dicionário de valências verbais do português brasileiro. Para isso foi criada uma notação das construções (diáteses), composta de uma análise sintática maximamente simples, mais os papéis semânticos de cada constituinte. Cada verbo é classificado de acordo com as diáteses em que pode ocorrer, o que fornece uma classificação detalhada, precisa e capaz de expressar adequadamente seu comportamento gramatical. Esse sistema substituiu com vantagem as classificações tradicionais atualmente correntes. São apresentados alguns problemas de análise ainda em discussão, tendo a ver principalmente com a definição e delimitação dos papéis semânticos. No que diz respeito à teoria linguística, a pesquisa tem demonstrado que a proporção geralmente pressuposta entre regularidade e irregularidade na estrutura das línguas subestima seriamente a importância da informação idiossincrática. Por exemplo, um levantamento piloto de 50 verbos já analisados no **Dicionário de valências**, escolhidos aleatoriamente, mostrou que se dividem em nada menos que 45 classes valenciais. Os resultados da pesquisa até o momento têm consequências para o planejamento de cursos do português língua estrangeira, e sugerem questões ligadas ao estudo da aquisição da linguagem.

Palavras-chave: Valências. Verbos. Gramática portuguesa. Lexicologia.

1. Introdução: o Dicionário de Valências do Português Brasileiro

O Projeto VVP (**Valências Verbais do Português Brasileiro**) tem como objetivo a elaboração de um dicionário de valências, onde cada verbo é apresentado juntamente com as construções em que pode ocorrer. O Projeto está sendo levado a efeito com a colaboração de pesquisadores de quatro universidades.¹ O objetivo é simples, mas a construção do Dicionário de Valências tem requerido a discussão e eventual esclarecimento de grande número de questões de teoria e análise. Neste

* Universidade Federal de Minas Gerais.

¹ UFMG (Mário A. Perini e Larissa Ciríaco); UFOP (Eliane Mourão); UFV (Aparecida de Oliveira); PUC-Minas (Madalena Loredó Neta). Artigo produzido em fevereiro de 2016.

artigo, apresento uma amostra dessas questões, com alguma discussão. Como se verá, algumas são relevantes para a análise da língua mesmo fora da área das valências verbais – em particular, na área da formulação das construções e na articulação entre os espaços sintático e semântico das estruturas da língua. As questões são numerosas, algumas muito complexas, e foram discutidas em textos maiores, principalmente em Perini (2015). Aqui seleciono alguns problemas para os quais já podemos apontar solução possível.

No momento, o **Dicionário** se compõe de cerca de 500 verbetes; o plano é incluir todos os verbos correntes no português brasileiro falado, cerca de 3.000 segundo nossa avaliação. Além dos verbetes, inclui certo número de miniartigos, que definem sumariamente alguns problemas relevantes encontrados durante a pesquisa, ou explicam particularidades da notação utilizada.

Os dados que fundamentam a pesquisa, e que aparecem em parte no **Dicionário**, são de proveniência variada: frases criadas, observadas no dia a dia, tiradas de *corpus* etc., sempre passando pelo filtro de nosso conhecimento da língua. Procuramos incluir apenas construções correntes no brasileiro falado, evitando formas puramente literárias, arcaicas e limitadas ao português europeu. Por exemplo, incluímos construções do tipo *meu carro furou o pneu*, mas não *furou-lhe as férias ser chamado para a tropa*². O **Dicionário** está sendo montado em sua versão em inglês, e pretende-se colocá-lo eventualmente na internet para consulta livre.

A motivação inicial de nossa pesquisa é a necessidade que vemos de elaborar descrições de amplas seções da estrutura da língua, como pré-requisito à construção de teorias. Seguimos aqui a orientação de Gross (1975), que observa que, na história das ciências,

Toda construção teórica foi sempre precedida de um longo trabalho de acumulação sistemática de dados, e os pesquisadores sempre se esforçaram por preencher as lacunas que podiam se apresentar em seus dados, antes de adiantar uma regra geral. (GROSS, 1975, p. 9)

É dentro dessa perspectiva que se coloca nosso trabalho no **Dicionário**. A bem da justiça, é bom observar que esse tipo de trabalho empírico de acumulação de dados sistematizados está em ascensão no momento; pode-se citar, primeiro, os diversos dicionários de valências ou obras assemelhadas que apareceram nos

² Busse (1994).

últimos anos, entre os quais ressaltamos dicionários de valências como o ADESSE (espanhol)³, DICOVALENCE (francês) e Herbst *et al.* (2004; inglês), assim como trabalhos de classificação de verbos de caráter mais geral, como Borba (1990), Levin (1993) e Busse (1994). Todos esses trabalhos⁴ se concentram no estudo empírico dos fatos da língua, minimizando os pressupostos teóricos.

O **Dicionário** tem, portanto, um objetivo diferente dos de muitos trabalhos atuais de teoria linguística: pretende apresentar os fatos relativos à valência verbal, utilizando para isso uma linguagem descritiva que é, na medida do possível, teoricamente neutra. Essa linguagem, entendida como um sistema notacional, é explicitada em Perini (2015); evidentemente, não é totalmente isenta de pressupostos teóricos, o que não seria mesmo possível. Mas houve a preocupação constante de usar a porção da teoria linguística que é de aceitação geral; essa porção existe, e não é nada pequena. Em alguns pontos, foi necessário introduzir noções teóricas pouco convencionais, que são todas explicadas no livro.

A dificuldade teórica maior ao construir o **Dicionário** é a de como exprimir as construções (diátases) que se associam a cada verbo. Para isso tivemos que elaborar um sistema específico, que esperamos será de fácil compreensão para os linguistas; o sistema é brevemente exposto na próxima seção. O resultado é a apresentação de uma base de dados analisados, cobrindo (eventualmente) a totalidade dos verbos do português. Dada a importância gramatical dos verbos e suas valências, o **Dicionário** representa a descrição de uma grande área dentro da estrutura da língua.

2 Notação

2.1 Traços básicos

A notação é uma versão simplificada da usada, por exemplo, por Goldberg (1995, 2006) e Jackendoff (1990). A simplificação se justifica porque o foco da descrição são as valências verbais: assim, detalhes irrelevantes para esse efeito (mas que podem ser importantes para a descrição de outros setores da estrutura) são deixados de lado. Basicamente, a notação de uma construção segue o modelo

3 O ADESSE é, de longe, o melhor de todos; foi elaborado sob a coordenação de José M. García-Miguel, da Universidade de Vigo.

4 E outros mais. Até o momento pudemos examinar dicionários de valências do francês, inglês, tcheco, espanhol e italiano.

seguinte: digamos que queremos descrever a possibilidade de o verbo *arranhar* ocorrer em frases como

[1] O gato arranhou minha canela.

A construção correspondente é notada como⁵

[2] SujV>Agente V SN>Paciente

que se pode ler como “sujeito Agente, verbo, SN Paciente”⁶.

O leitor observará que o segundo SN não é mencionado por sua função sintática (“objeto direto”). Isso é um aspecto da análise sintática que baseia a notação. Enquanto os trabalhos sobre valência em geral simplesmente adotam a análise sintática tradicional, com seu elenco de funções (sujeito, objeto direto, predicativo, objeto indireto, adjunto adverbial etc.), o **Dicionário** utiliza uma análise muito mais simples, que apenas distingue uma função sintática (sujeito), sendo os outros termos referidos apenas por sua classe (SN, SAdv, V...), a eventual preposição que os introduz (notada especificamente, ou seja, *de*, *com* etc.) e sua ordem na sequência. Essa análise é bem fundamentada no que diz respeito à descrição das valências; foi adotada não por ser simples, mas porque descreve adequadamente todos os aspectos sintáticos da estrutura relevantes para a descrição das valências.

Além da sequência sintática (em [2], **SujV V SN**), a notação inclui os papéis semânticos de cada um dos constituintes relevantes (em [2], Agente para o sujeito, Paciente para o outro SN). Nada mais é necessário para subclassificar os verbos quanto a sua valência. O verbo *arranhar*, portanto, é descrito como podendo ocorrer em certo conjunto de construções, como [2] e outras. Esse conjunto de construções em que um verbo pode ocorrer é sua **valência**. Desse modo, o verbete *arranhar* aparece assim no **Dicionário**⁷:

5 Na verdade, a fórmula usa **VSubj, Agent, NP, Patient**, porque o texto do **Dicionário**, em sua versão atual, está em inglês. **VSubj**, ou **SujV** (sujeito valencial), pode ser entendido simplesmente como o sujeito; a razão do restritivo *valencial* é dada em Perini (2015, p. 9), mas não precisa ser explicada aqui.

6 Adotamos a convenção de grafar os papéis semânticos com maiúscula.

7 Sintagmas em *itálico* não fazem parte da construção; ver 2.4 abaixo. Os códigos ‘C1’, ‘C4’ etc. servem para facilitar a referência às diáteses. Cada diátese do **Dicionário** tem um código formado da letra C, mais um número arbitrário.

ARRANHAR

C1 **SujV**>Agente **V** **SN**>Paciente

Esse gato arranhou minha perna.

.....
C2 **SujV**>Agente **V**

Esse gato arranha.

.....
C3 **SujV**>Paciente **Refl** **V**

O menino se arranhou *todo*.

.....
C4 **SujV**>Paciente **V**

O menino arranhou *todo*.

Quando um termo é preposicionado, menciona-se a preposição particular que o introduz; assim, as frases

[3] Eu contava com vocês.

[4] O canalha bateu no meu cachorro.

se analisam, respectivamente, como

C61 **SujV**>*Experienciador* **V** *com* **SN**>*Estímulo*

C79 **SujV**>Agente **V** *em* **SN**>Paciente

Essa notação especificada é necessária, em vez de usar o termo *sintagma preposicionado*, porque a identidade da preposição é parte da valência do verbo: com *contar*, temos *com*, mas com *bater* ocorre *em*, com *gostar* ocorre *de* etc.

2.2. Construções complexas

Um tipo de construção que ocorre com diversos verbos do **Dicionário**, mas que não é de uso frequente, são as *construções complexas*. Em alguns casos uma oração simples precisa ser semanticamente analisada como um conjunto de

proposições. Tais casos foram apontados já por Jackendoff (1972, p. 35). Para pegar um exemplo do português, temos a frase:

[5] A Sara vendeu um lote para o Vítor por 80 mil reais.

Quando vamos anotar essa estrutura quanto aos papéis semânticos expressos, chegamos a um aparente impasse: por exemplo, *Sara*, além de ser o Agente, é a Fonte (pois o lote passa da posse de Sara para a de Vítor), mas também a Meta (do dinheiro, que passa de Vítor para Sara). Ou seja, para representar corretamente a semântica da oração, não basta marcar *Sara* como simultaneamente Fonte e Meta: ela é a Fonte do lote e a Meta do dinheiro. A transferência dupla de posses é essencial no significado e na valência de *vender*, que se distingue de *dar* exatamente por isso.

A solução adotada no *Dicionário* é a de distinguir dois eventos, simultâneos e interdependentes, em uma representação que denominamos *construção complexa*; a construção correspondente a [5] é a seguinte:

C74 Sintaxe: SujV	V	SN	para SN	por SN
Ev.1 <i>Agente+Fonte</i>		<i>Tema</i>	<i>Meta</i>	
Ev.2 <i>Meta</i>			<i>Fonte</i>	<i>Tema</i>
a Sara	vendeu	um lote	para o Vítor	por 80 mil reais

Como se vê, cada constituinte recebe pelo menos um papel semântico, mas o Tema do primeiro evento é *o lote*, e o Tema do segundo evento é *80 mil reais*. Dessa forma, fica representada a dualidade de transferências de posse que o verbo *vender* necessariamente expressa.

Há um bom número de verbos que ocorrem em construções complexas: *vender*, *comprar*, *pedir*, *trocar*, *impedir*, *permitir* e outros. Em todos esses casos é necessário distinguir mais de um evento para que se possa atribuir os papéis semânticos aos respectivos constituintes de modo intuitivamente satisfatório.

Os eventos não precisam ser sempre entendidos como factuais. Em [5] entende-se que as duas transferências foram realizadas; mas em

[6] Robinho pediu dinheiro ao pai.

temos um evento factual (Robinho pedir alguma coisa, uma ação *dicendi*) e um segundo evento, o pai dar dinheiro a Robinho, do qual só se asseve o *desejo* de Robinho de que se realize; a frase deixa em aberto se o pai efetivamente deu o dinheiro. Desse modo, um elemento a ser notado na construção é a relação entre os eventos: para [5], ambos são asseridos como ocorrendo, mas para [6] um é asserido como ocorrendo, e o outro como desejado. Note-se como esses detalhes também são parte integrante do significado e da valência do verbo da oração.

2.3. Valência e a classificação dos verbos

Um verbo pode ocorrer em bom número de construções (o mais complexo que encontramos até agora é *passar*, que aparece no **Dicionário** em nada menos que 23 construções). Mas há ainda outras construções, que não aparecem no **Dicionário** porque não são lexicalmente condicionadas e não são consideradas parte da valência dos verbos respectivos.

Para entender a diferença entre construções que aparecem na valência (e que se denominam **diáteses**) e construções que não aparecem, vamos pegar um exemplo. Seja a construção negativa, que aparece em [7] A inundação não chegou até a minha casa.

Essa frase realiza uma construção diferente de [8] A inundação chegou até a minha casa.

Formalmente, [7] tem a partícula *não*, ausente de [8]; e semanticamente [7] nega o que [8] afirma. Logo, são duas construções diferentes. Mas ocorrência na construção negativa não é uma característica do verbo *chegar*: na verdade, qualquer verbo da língua pode ser negado, exatamente da mesma forma e com exatamente o mesmo efeito semântico. Ou seja, o que temos aqui é um fenômeno geral (que pode ser classificado como gramatical) e não um fenômeno individualizado (lexical), e não faria sentido incluir a construção negativa na valência do verbo *chegar*, ou de qualquer verbo.

Essa situação contrasta com a da construção [9] A toalha secou.

Aqui temos uma construção onde o sujeito é Paciente, ao contrário de

[10] A m secou a toalha.

na qual o sujeito é Agente e o objeto (*a toalha*) é Paciente. Mas a construção ilustrada em [9] não ocorre com qualquer verbo, mesmo se só considerarmos os verbos que aceitam um Paciente. Assim, não podemos dizer

[11] * A pizza comeu.

com sujeito Paciente, correspondente a

[12] O menino comeu a pizza.

Ou seja, ocorrência na construção [9] caracteriza o verbo *secar* em oposição a *comer*. Em outras palavras, a construção ilustrada em [9] divide os verbos da língua em dois grupos: os que podem e os que não podem ocorrer nela; esse tipo de construção é denominado *diátese*: a construção exemplificada em [9] é uma das diáteses do verbo *secar*. Já a construção negativa, como é claro, não subclassifica os verbos, já que todos eles podem ser negados; assim, essa construção não é uma diátese.

A *valência* de um verbo é o conjunto de suas diáteses; o verbo *passar*, como vimos, tem uma valência muito complexa, compreendendo (pelo menos) 23 diáteses. Já *acariciar* tem uma valência simples, e só pode ocorrer em uma diátese. **O Dicionário de valências verbais do português brasileiro** é a lista de todos os verbos da língua, cada um com sua valência. Os verbos se dividem em classes valenciais e cada classe contém os verbos de valência idêntica.

2.4. “Complementos” e “adjuntos”

Um detalhe da formulação das diáteses tem levantado bastante discussão: o que, exatamente, devemos colocar na fórmula, e o que pode ficar de fora? Ou seja, podemos ou não podemos considerar as duas frases abaixo como realizações da mesma diátese de *secar*?

[10] A máquina secou a toalha.

[13] A máquina secou a toalha em menos de cinco minutos.

O constituinte *em menos de cinco minutos* indica duração. Mas esse valor (relação semântica) não depende da valência do verbo; o sintagma *em menos de cinco minutos* só pode significar um período de tempo, independentemente da função sintática que tenha, e independentemente do verbo da oração. Compare-se com *a máquina*, que pode ser Agente, como em [13], ou Paciente, como em

[14] Ronaldo consertou a máquina.

Nessa construção *a máquina* é Agente quando é sujeito e Paciente quando é objeto – ou seja, seu papel semântico depende da função, e, portanto, esse constituinte precisa aparecer na diátese. Esse é o critério básico que determina a inclusão de constituintes na diátese: registram-se os constituintes cujo papel semântico depende de sua função na construção. Essa é informação relacionada com o verbo da oração: com *consertar*, o sujeito é Agente, mas com *receber* o sujeito é a Meta, ou seja, o final de um movimento real ou virtual:

[15] Eu recebi uma encomenda pelo correio.

Outro fator que determina a ocorrência na diátese é a ocorrência obrigatória. Esse fenômeno é relativamente raro, mas aparece às vezes, e também precisa ser atribuído às propriedades do verbo. Assim, um sintagma locativo como *em Fortaleza* é de ocorrência opcional em

[16] Hoje choveu em Fortaleza.

mas é obrigatório em

[17] Ela morava em Fortaleza.⁸

Por isso o complemento de Lugar aparece na diátese correspondente de *morar*, mas não na de *chover*.

⁸ Com *morar*, é obrigatório um complemento de Lugar (como em [17]), ou de Companhia, ou ainda de Modo; mas *morar* não ocorre sem complemento.

Esse critério de dependência do verbo é que determina a inclusão ou não de um elemento na diátese; e ele substitui a oposição tradicional entre “complementos” e “adjuntos” – em nossa nomenclatura, todos são chamados **complementos**. Ou seja, para responder a uma pergunta anterior, as frases

[10] A máquina secou a toalha.

[13] A máquina secou a toalha em menos de cinco minutos.

são realizações da mesma diátese, que inclui um sujeito Agente e um objeto Paciente. O constituinte *em menos de cinco minutos* não se relaciona com as propriedades sintáticas e semânticas do verbo: tem forma e papel semântico autônomos.

Essas são as características básicas na notação utilizada no **Dicionário**. Além delas, há certo número de detalhes que não precisam ser apresentados aqui; o leitor interessado pode consultar Perini (2008; 2015) para uma explanação completa.

3. Valências e a classificação dos verbos

3.1. Classes de verbos no Dicionário

Uma vez definidas as diáteses da maneira esboçada acima, cada verbo é classificado segundo as diáteses em que ocorre, ou seja, segundo sua valência. A valência descreve com precisão o comportamento gramatical de um verbo: que complementos co-ocorrem com ele, e que papéis semânticos cada um desses complementos veicula. Para dar um exemplo, peguemos o verbo *chorar*. Esse verbo ocorre em duas diáteses, exemplificadas por:

[18] O bebê está chorando.

[19] Ela chorou a morte do pai.

Em [18] temos sujeito Agente, e nada mais; em [19] temos sujeito Agente e objeto Causa. No **Dicionário**, isso se representa da seguinte maneira:

CHORAR

C2 SujV>Agente

O bebê está chorando.

C106 SujV>Agente V SN>Causa

Ela chorou a morte do pai.

Agora, compare-se *chorar* com *comer*: este último verbo ocorre em C2, mas não ocorre em C106, porque embora possa aparecer com sujeito e objeto, o objeto é Paciente e não Causa (esta é a diátese C1). Já *acariciar* ocorre em C1, mas não em C2 (sem objeto), nem em C106 (com objeto Causa). E *ser* não ocorre em nenhuma das diáteses vistas, e na verdade nunca comporta Agente nem Paciente; em vez disso, aparece (entre outras) em uma diátese que simplesmente assere a identidade referencial de duas entidades mencionadas:

[20] Aquela senhora loura é a professora de História.

Comer, chorar, acariciar nunca ocorrem nessa diátese. Com esses verbos, o sujeito e o objeto claramente denotam entidades de referência distinta: por exemplo, a frase

[10] A máquina secou a toalha.

deixa claro que a máquina é uma coisa e a toalha é outra; mas [20] assere que a senhora loura e a professora de História são a mesma pessoa.

Pode também haver diferenças formais: em [10] o Agente é o sujeito, e o Paciente é outro SN. Mas em:

[21] A máquina acabou com a toalha.

o Paciente é sinalizado por um sintagma preposicionado. Cada uma dessas diferenças, de estrutura sintática e/ou de papel semântico, determina uma nova diátese.

3.2. Valências e a classificação tradicional

Esse sistema descreve o comportamento gramatical dos verbos de maneira extremamente detalhada e precisa. Para se ter uma ideia, temos no momento cerca de 500 verbos no **Dicionário**, que se dividem em mais de 100 classes, sendo cada classe o conjunto dos verbos de valência idêntica. Um levantamento piloto de 50 verbos, escolhidos aleatoriamente no **Dicionário**, mostrou uma divisão em nada menos que 45 classes distintas⁹. Isso mostra claramente a necessidade de mapear o léxico, antes de partir para grandes generalizações.

As diáteses definidas até o momento são mais de 200, e certamente ainda há outras a serem definidas. É preciso enfatizar que o conhecimento de todo esse imenso sistema é parte integrante do conhecimento da língua: nenhum falante diz **eu gosto pizza*, nem **a pizza comeu em cinco minutos* (embora diga *a pizza acabou em cinco minutos*) etc., e tudo isso depende crucialmente do conhecimento das valências dos verbos respectivos. Um falante conhece milhares de verbos, e aplica com toda segurança as valências de todos eles. Esses resultados são muito significativos, e mostram que a complexidade da estrutura da língua é muito maior do que às vezes se imagina.

É instrutivo comparar esse sistema com o da gramática tradicional. Este define (em sua versão mais usual) apenas cinco classes de verbos: *intransitivo*, *transitivo direto*, *transitivo indireto*, *transitivo direto e indireto* e *de ligação* (cf. HAUY, 2014, pp. 1072-1103). É fácil verificar que esse sistema tão simples não dá conta de descrever as diferenças entre os verbos; para dar apenas uns poucos exemplos:

Comer ocorre com ou sem objeto; é “transitivo” ou “intransitivo”?

Dizemos *contou com você*, mas *gosto de você*; no entanto, apesar da diferença de preposições, ambos esses verbos precisam ser analisados como “transitivos indiretos”.

Morrer e *pular* ambos podem ocorrer sem objeto: *o gato morreu*, *o gato pulou*. Seriam, presumivelmente, “intransitivos”. Mas em um caso o sujeito é Paciente, e no outro é Agente – essa diferença, essencial para o uso desses verbos, não pode ser expressa no sistema tradicional.

O objeto direto tradicional expressa uma grande variedade de papéis semânticos: Paciente em *Fred comeu a pizza*; Causa em *Fred chorou a morte*

⁹ Ver alguns exemplos no Apêndice.

do pai; Estímulo (causador de experiência) em *Fred sentiu uma pontada*; Coisa qualificada em *eu acho a sua filha linda*; Meta (final de um movimento) em *os alpinistas atingiram o pico* etc. O conhecimento desse fato é essencial para o uso correto desses verbos, mas não há como representá-lo no sistema tradicional.

O sistema tradicional, é verdade, é muito mais simples do que o apresentado no **Dicionário**. No entanto, não chega nem perto de descrever os dados adequadamente – além da dificuldade de ser aplicado, devida principalmente à falta de definições objetivas das unidades e funções sintáticas e semânticas utilizadas. Um sistema complexo não pode ser descrito de forma simples.

Para avaliar a pouca capacidade do sistema tradicional de expressar os fatos gramaticais (sintáticos e principalmente semânticos), basta examinar a lista seguinte, com frases que são todas analisadas tradicionalmente como compostas de **sujeito, verbo e objeto direto**:

- O Fido comeu a banana. [Agente – Paciente]
- A multidão deixou o estádio. [Tema – Fonte]
- Os alpinistas atingiram o pico. [Tema – Meta]
- Todo mundo adorou a festa. [Experienciador – Estímulo]
- A proposta agradou o chefe. [Estímulo – Experienciador]
- Meu primo tem dois carros. [Possuidor – Coisa.possuída]
- Ela pesava cem quilos. [Coisa.localizada – Medida]
- O político aceitou o dinheiro. [Meta+Agente – Tema]
- O político recebeu um tapa. [Meta (*não* Agente) – Tema]

Temos aqui dez situações simbolicamente bem diferentes – ou seja, dez construções sintaticamente idênticas, mas muito distintas quanto ao que expressam. Chamar todos esses verbos (e/ou construções) de “transitivos” é escamotear a maior parte de seu significado gramatical.

4. Resultados e problemas

A dificuldade principal encontrada na construção do *Dicionário* é como formular as diáteses; e a maior dificuldade dessa formulação é como identificar e delimitar os papéis semânticos. Por outro lado, não é possível esperar até que se disponha de uma lista razoavelmente segura de papéis semânticos para começar a

construir o **Dicionário**, porque a definição e a delimitação dos papéis dependem, por sua vez, do estudo de muitos exemplos. Ou seja, construir o **Dicionário** depende da definição dos papéis semânticos, e a definição dos papéis depende da lista de verbos e valências. A saída desse dilema é, claramente, trabalhar simultaneamente nas duas frentes. Por conseguinte tanto o **Dicionário** quanto os papéis semânticos devem ser entendidos como propostas temporárias, em revisão constante: cada verbete se entende como resposta a algumas perguntas, e também – muitas vezes principalmente – como o levantamento de novas questões.

4.1. Análise sintática

A análise sintática utilizada, e que foi resumida na seção 2.1 acima, é muito simples, e tem sido suficiente para as necessidades da descrição das diáteses. Não sabemos se a descrição de outras áreas da estrutura levará à introdução de complicações. Por ora, podemos satisfazer-nos com o seguinte:

- uma sequência de símbolos categoriais: SN, SAdj, SAdv;
- preposições individualmente notadas: *com, em, de, a, para e por* (todas as outras preposições ocorrem com constituintes não valenciais);
- uma função sintática, a saber, o sujeito.

Isso é tudo. As outras diferenças, que são muitas, são descritas em termos de papéis e outras relações semânticas. Assim, pode-se mostrar que o chamado “predicativo” é apenas um sintagma (adjetivo ou nominal) com um papel semântico qualificativo – e, portanto, emparelhado semanticamente com outro sintagma, que expressa a coisa qualificada, como em

[22] Esse menino é uma peste.

[22] se analisa sintaticamente como **SujV V SN**, ou seja, de maneira idêntica a

[23] Esse menino quebrou a janela.

A diferença está em que em [22] o sujeito é Coisa qualificada e o outro SN Qualidade; em [23] o sujeito é Agente e o SN não sujeito Paciente. Isso vem naturalmente da semântica do verbo, que não precisa ser repetida nas funções sintáticas.

4.2. Papéis semânticos

O que falta em complicação na sintaxe sobra na semântica – em especial na definição e delimitação dos papéis semânticos. A lista tradicional – Agente, Paciente, Instrumento, Lugar, Fonte, Tema etc. – não cobre todos os casos. E, o que é pior, há um número excessivo de casos em que não se sabe direito que papel devemos colocar, porque eles não estão claramente delimitados. Por exemplo, na frase

[24] O cachorro correu para o quintal.

o sujeito é certamente um Agente, pois pratica a ação denotada; mas é também o Tema, ou seja, o elemento que se desloca. Não é claro, de imediato, qual desses papéis se deve usar na definição da construção, ou se devemos acumular os dois, o que também é possível: o sujeito seria Agente + Tema¹⁰.

O Paciente é particularmente problemático. Entre os termos tradicionalmente analisados como Paciente, podem-se enumerar as seguintes relações semânticas:

“mudança de estrutura ou de aparência”

[25] Quebrei o jarro.

“início de existência”

[26] Compus uma canção.

“desaparecimento”

[27] A gasolina acabou.

[28] O governo extinguiu a agência.

“tema”

[29] Arrastei a mesa para o quintal.

[24] O cachorro correu para o quintal.

¹⁰ A acumulação de papéis semânticos já é utilizada em Jackendoff (1972). Na verdade, é um artefato de nossa ignorância atual sobre a definição dessas relações.

e ainda outros, incluindo alguns difíceis de definir, como
[30] Aplaudi a decisão.

[31] Os incas adoravam o sol.

[32] Achei uma carteira na rua.

O que é, exatamente, um Paciente? Aqui precisamos de uma relação semântica coerente, codificada coerentemente pela sintaxe; até o momento não a encontramos, de modo que qualquer diátese que inclua um Paciente está em discussão.

Assim como o Paciente, há outros papéis semânticos que ainda esperam uma definição clara e operacional. Esse é um dos problemas principais que nos ocupam no momento em nosso esforço de construir o **Dicionário**. É bom observar que o problema não foi satisfatoriamente solucionado por nenhum dos dicionários e bases de dados que pudemos examinar: a maioria se contenta com a lista tradicional, ou com soluções *ad hoc* para cada caso particular. Por outro lado, tem havido progresso, e se encontra na literatura alguma discussão relevante sobre o assunto: posso citar em especial os trabalhos de Dowty (1989; 1991) e Cançado (2003), que propõem uma análise dos papéis semânticos em traços tais como <afetado>, <estativo> etc. Isso pode representar uma base para desenvolver uma compreensão melhor do fenômeno; mas verificar e aplicar essas hipóteses aos exemplos disponíveis é ainda uma tarefa futura.

5. O que o *Dicionário* pode nos ensinar

O *Dicionário*, mesmo em sua forma atual, um tanto preliminar, já mostra vários fatos interessantes a respeito da língua. Em primeiro lugar, fica evidente que a proporção geralmente pressuposta entre regularidade e idiosincrasia subestima seriamente a importância desta última. Com efeito, a quantidade de informação vinculada a itens lexicais, e não redutível a regra geral, é imensa; basta lembrar os 50 verbos examinados, que se dividem em 45 classes valenciais. Examinando os verbos um a um encontramos diáteses como a que chamamos C1, a saber,

C1 SujV>Agente V SN>Paciente

Essa diátese está presente na valência de verbos tão variados como *comer, quebrar, dobrar, pintar, destruir, chutar, beijar* e mais centenas, talvez milhares de verbos. Por outro lado, a diátese

C279 SujV>Agente V de SN>Paciente SN>Qualidade

só ocorre com um verbo da língua, *fazer*, como em

[33] O governo fez da crise uma tragédia.

Esse fato nos leva a repensar o próprio processo de aprendizagem da língua, pois requer uma capacidade maior do que geralmente se imagina de aprender elementos individuais, não conectados por regras gerais. Isso tem consequências para o planejamento do ensino do português como língua estrangeira, pois leva necessariamente a uma ênfase maior nos detalhes das propriedades de itens lexicais específicos. Em outras palavras, a estrutura da língua é vastamente mais complexa do que sugere pelo menos uma parte das teorias correntes.

Ainda no que diz respeito à aprendizagem da língua, o grande número de classes de verbos sugere que a aprendizagem se dirige não à classe de cada verbo, mas às diáteses que o classificam. Em outras palavras, nossa mente ao encontrar um novo verbo não se pergunta “é transitivo, intransitivo ou de ligação?”, mas “em que diáteses esse verbo pode ocorrer?”. Fica assim explicado o fenômeno da classificação equivocada como quando uma criança diz *você bateu o cachorro* – ela está atribuindo ao verbo *bater*, no sentido de ‘espancar’, uma diátese que se aplica à grande maioria dos verbos de ação (ou seja, C1). Só mais tarde ela vai aprender que esse verbo é excepcional, e exprime Agente e Paciente através de uma diátese idiossincrática, com a preposição *em* antes do Paciente.

Finalmente, a construção do **Dicionário** vem mostrando com muita clareza um ponto bastante citado, mas raramente exemplificado, a saber, o caráter contínuo da divisão entre gramática e léxico. Ao examinar as valências, vemos com clareza que isso é verdadeiro, e também podemos dar exemplos. Assim, a associação entre sujeito e Agente (que se observa nas diáteses C1, C2, C74, C79, C106, todas vistas acima) aparece em dezenas de diáteses, e pode ser considerada resultado da

aplicação de uma regra gramatical de encadeamento (*linking rule*), que comporta exceções, mas se aplica à imensa maioria dos casos. Mas a associação do Agente com um sintagma da forma *de SN* ocorre apenas com um verbo da língua, a saber

[34] O vizinho apanhava da mulher.

e precisa ser marcada como um fenômeno lexical. Temos aqui o mesmo processo (associação entre função sintática e papel semântico) sendo forçosamente descrito em termos gramaticais em um caso, em termos lexicais em outro – um exemplo da falta de distinção nítida entre gramática e léxico.

Building the Valency Dictionary: problems and results.

Abstract

Project VVP (in course at UFMG with participation of researchers from three other Universities of Minas Gerais) has as its aim the elaboration of a dictionary of verb valencies of Brazilian Portuguese. A notation is proposed for constructions (**diatheses**), which utilizes a maximally simple syntactic analysis, complemented by the semantic roles assigned to each constituent. Each verb is thus classified according to the diatheses it can occur in, providing a detailed and precise classification, which adequately expresses the grammatical behavior of each verb. This system replaces with advantage current traditional classifications. Some problems under discussion at the moment are presented, which have to do with the definition and delimitation of semantic roles. In what regards linguistic theory, the research has shown that the usually assumed proportion between regular and irregular phenomena in the structure of languages greatly underestimates the importance of idiosyncratic information. For instance, a preliminary survey of 50 verbs in the *Valency dictionary*, chosen randomly, showed that they belong to no less than 45 valential classes. The results of the research so far are relevant for the planning of courses in Portuguese as a foreign language, and suggest questions having to do with the study of language acquisition.

Keywords: Valency. Verb. Portuguese grammar. Lexicology

Referências

ADESSE (**Alternancias de diátesis y esquemas sintáctico-semánticos del español**). Disponível em <<http://adesse.uvigo.es>>

BORBA, Francisco da Silva (Ed.) **Dicionário gramatical de verbos do português contemporâneo do Brasil**. São Paulo: Ed. UNESP, 1990.

BUSSE, Winfried (org.). **Dicionário sintáctico de verbos portugueses**. Coimbra: Almedina, 1994.

CANÇADO, Márcia. “Um estatuto teórico para os papéis temáticos”. In: **Semântica Formal**. MÜLLER, Ana Lúcia; NEGRÃO, Esmeralda Negrão Vailati; FOLTRAN, Maria José (Org.) São Paulo: Contexto, 2003.

CHIERCHIA, Gennaro, Barbara H. PARTEE e R. TURNER (org.). **Properties, types and meaning**. Studies in Linguistic and Philosophy, vol. II: Semantic Issues. Dordrecht: Kluwer, 1989.

CULICOVER, Peter W. e Ray S. JACKENDOFF. **Simpler Syntax**. Oxford: Oxford University Press, 2005.

DICOVALENCE 2. **Dictionnaire de valence des verbes français**. 2010. Disponível em <bach.arts.kuleuven.be/dicovalence/>

DOWTY, David. “On the semantic content of the notion of ‘thematic role’”. In: CHIERCHIA, G.; PARTEE, B.; TURNER, R. (Ed.). **Properties, Types and Meaning**. Dordrecht: Kluwer, 1989. p.69-129.

DOWTY, David. “Thematic proto-roles and argument selection”. **Language**, 67, 3, 1991.

GOLDBERG, Adele. **Constructions at work**. Oxford University Press, 2006.

GOLDBERG, Adele. **Constructions – a construction grammar approach to argument structure**. The University of Chicago Press, 1995.

GROSS, Maurice. **Méthodes en syntaxe**. Paris: Hermann, 1975.

HAUY, Amini B. **Gramática da língua portuguesa padrão**. São Paulo: EDUSP, 2014.

HERBST, Thomas, HEATH, David; ROE, Ian F.; GÖTZ, Dieter. **A valency dictionary of English: a corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives**. Berlin: Mouton de Gruyter, 2004.

JACKENDOFF, Ray S. **Semantic interpretation in generative grammar**. Cambridge, MA: MIT Press, 1972.

JACKENDOFF, Ray S. **Semantic structures**. Cambridge, MA: MIT Press, 1990.

LEVIN, Beth. **English verb classes and alternations**. University of Chicago Press, 1993.

MÜLLER, Ana Lúcia; NEGRÃO, Esmeralda Negrão; FOLTRAN, Maria José (Org.) **Semântica Formal**. São Paulo: Contexto, 2003.

PERINI, Mário A. **Estudos de gramática descritiva: as valências verbais**. São Paulo: Parábola, 2008.

PERINI, Mário A. **Describing verb valency: practical and theoretical issues**. Cham, Switzerland: Springer, 2015.

Apêndice: classificando verbos

Dou a seguir uma lista de alguns verbos, devidamente classificados quanto a sua valência, segundo a notação utilizada no *Dicionário*:

C1: *acariciar, assaltar, derrotar, devorar*

C1, C4: *amassar, engordar, adormecer, afundar*

C1, C2: *aplaudir, atacar, babar, beliscar, cantar*

C1, como vimos, é a diátese que tem sujeito Agente e mais um SN Paciente. C4 tem sujeito Paciente e não comporta outro SN; por exemplo, *a lata amassou*. C2 tem sujeito Agente e não tem expressão do Paciente: *todos aplaudiram*. Como vemos, temos aí verbos que só ocorrem em C1, outros que ocorrem em C1 e C4 e ainda outros que ocorrem em C1 e C2.

Já outros verbos têm valência muito mais complexa. Sem entrar em detalhes, basta examinar os seguintes:

C1, C4, C16, C79, C86, C88: *bater*

C23, C54, C100: *chegar*

C1, C4, C43, C46, C47, C48, C102: *acabar*

C1, C4, C16, C18, C22, C23, C28, C38, C97, C100, C105, C130, C138, C143,

C145, C174, C215, C216, C229, C230, C251, C252, C270: *passar*

Como já vimos acima, *passar* é até agora o verbo de valência mais rica que encontramos.

Submetido: 12/02/2016

Aceite: 08/07/2016